Question 1: Dental health in Singapore

Staff at Singapore Dental Hospital carried out a study to examine the dental health of 5-year-old children living in Singapore. Each of 360 children had a dental examination under standard conditions and the number of decayed, missing and filled teeth was recorded. Also recorded were the sex of the child, whether they were currently registered at a dentist and the deprivation category of the area (postcode sector, such as S12 8, or S25 7) the child lived in. The researchers’ question of interest was whether 5-year-old children living in deprived areas have worse oral health than 5-year-old children living in more affluent areas.

You have been given a subset of 100 of these children in the dataset. The data is stored as comma separated values (csv), a type to text file – Minitab, SPSS and Excel can open these straightforwardly. If opening in SPSS, it may ask you a few questions about the format. You can bypass these by also saving the format file **CSVfileFormat.tpf** and when opening your csv file, choose the tpf file when SPSS asks “Does your text file match a predefined format?” Alternatively, open the csv file in Excel and save it as a .xlsx file.

Variables in your data file:

**RefNumber** Child reference number (for admin purposes)

**DepCat** Deprivation category (on a 1-7 scale, where 1 is most affluent

and 7 is most deprived). In the catchment area for this study,

all the postcode sectors were in DepCat 4, 6 or 7

**RegCat** String (character) variable, with 3 categories:

NotReg – never registered with a dentist;

Lapsed – previously registered but now lapsed;

Reg – currently registered with a dentist

**Sex** 1 = boys, 2 = girls

**DFMT** Number of teeth with active decay, that are filled or are missing

a) Using either Minitab or SPSS, obtain appropriate descriptive statistics and suitable plots for the variables Sex, DepCat, RegCat and DFMT individually (i.e. **do not** look at any relationships between these variables). Provide a short interpretation of the output you produce for each variable.

b) Using either Minitab or SPSS carry out a 2-sample t-test and confidence interval to investigate whether there is a difference between oral health, as measured by DFMT, in the poorest areas (DepCat 7) and in less deprived areas.

HINT: You need to create a new variable (Poorest) that takes the value 1 if DepCat is 7 and 0 if it is less than 7. The categories of this new binary variable can be thought of as “lives in one of the poorest areas” (when Poorest=1) and “not living in one of the poorest areas” (when Poorest =0). You can use **Data > Code > Numeric to Numeric** to create the variable in Minitab and **Transform > Recode into Different Variables** to create the variable in SPSS.

Question 2:

State the assumptions of regression that can be assessed using residual plots. For each of these, state which residual plot or plots are useful for assessing the assumption and describe how the assumption can be checked, giving brief details of what should be looked for in each plot. [Note: there is no need to analyze any data in this question.]